



# Psicometría

## Tema 8 ANÁLISIS DE LA CALIDAD MÉTRICA DE LOS ÍTEMS

Psicometría

**M<sup>a</sup> Isabel García Barbero** , UNED. Universidad  
Nacional de Educación a Distancia, 2006

Esperanza Sánchez Moya



- La evaluación de la calidad métrica de los elementos del test nos informa de las características estadísticas de estos y de su contribución a la medición del constructo
- Los ítems deben ser capaces de discriminar entre sujetos
- Las características psicométricas de los ítems contribuyen a las características del test: su fiabilidad y validez.
- El análisis se lleva a cabo en dos fases: en el estudio piloto y en la confección de la prueba final
- El análisis cuantitativo de los ítems consiste en:
  - ✓ **Estimar sus parámetros:** dificultad, discriminación, fiabilidad y validez
  - ✓ **Análisis de los distractores**
  - ✓ **Sesgo:** Funcionamiento diferencial de los ítems (FDI)



# Propiedades psicométricas de los ítems

## I. La dificultad de un ítem

Aspecto que se establece a priori desde consideraciones teóricas y que están en función del contenido.

## II. La discriminación del ítem

Un ítem tiene poder discriminativo cuando es capaz de diferenciar adecuadamente a los sujetos de diferentes niveles de una variable criterio

## III. La fiabilidad y la validez de un ítems

Relacionadas con la discriminación de los ítems

## IV. Comparación de las respuestas de los ítems

Medir el grado de acuerdo u homogeneidad entre los ítems, ver si los ítems miden la misma variable

## V. Análisis de distractores

## VI. Funcionamiento Diferencial del Ítem

Analiza si distintos grupos con el mismo nivel de competencia tienen distintas probabilidades de éxito en un ítem, pero sin apuntar a posibles causas



# I. Dificultad del Ítem

- Solo para tests que miden variables cognitivas, en los que existen respuestas correctas e incorrectas

$$P = \frac{A}{N}$$

**A** = numero de sujetos que responden correctamente al ítem

**N** = numero de sujetos que responden al ítem

- Oscila entre 0 y 1
- Ítems con valores extremos  $P=0$   $P=1$  deben ser descartados
- **Para ítems dicotómicos**

$$\mu = P$$

$$\sigma^2 = P(1 - P) = P * Q$$

- Se debería de llamar **índice de facilidad** ya que los valores altos indican la facilidad del ítems
- Su principal inconveniente es que depende de la muestra de sujetos en que se calcule, de tal forma que un mismo ítem tendrá tantos índices de dificultad como muestras se puedan utilizar para su cálculo.



- **Corrección de la dificultad en ítems de elección múltiple**

- ✓ Se aplica cuando sospechamos que existe un subgrupo de sujetos relativamente numeroso que responden correctamente por adivinación y que para este grupo todas las alternativas tienen la misma probabilidad de ser elegidas

$$P = \frac{A - \frac{E}{k-1}}{N}$$

- ✓ La dificultad de los ítems afecta claramente a la variabilidad de las puntuaciones en el test y a la precisión con la que la puntuaciones del test discrimina entre diferentes grupos de sujetos



## Ejemplo

En la tabla adjunta aparecen las puntuaciones obtenidas por 10 sujetos en un test X que consta de 15 ítems.

Calcular el **ID** de los ítems y ver si algunos deberían de ser eliminados.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0	1	1	1	1	1	0	1	0	1	1	0	0	0	0
2	1	1	0	0	1	1	1	1	1	1	0	1	0	0	1
3	1	0	0	0	1	1	0	0	1	1	1	1	1	0	1
4	1	1	0	0	1	1	0	0	1	1	0	1	1	0	1
5	0	0	0	0	1	0	0	1	0	1	0	0	1	0	1
6	1	1	0	0	1	1	0	0	0	0	0	0	0	0	1
7	1	1	1	1	1	1	0	0	0	0	1	1	1	0	1
8	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
10	1	0	1	0	1	1	0	1	0	1	0	1	0	0	1



	A	B	C	D
20% superior	8	17	207	11
20% Siguiete	11	39	223	25
20% Siguiete	16	53	183	32
20% Siguiete	17	61	113	29
20% Inferior	43	70	106	39
p	0.07	0.18	0.64	0.10

### Ejemplo:

En la tabla aparecen las frecuencias de sujetos, ordenados en 5 percentiles según su puntuación, que respondieron a un ítem de un test de rendimiento en lengua con 4 opciones de respuesta, donde la opción C es la correcta. La media y varianza en el test de los sujetos fue de 6 y 4 respectivamente, mientras que la media en el test de los sujetos que acertaron el ítem fue de 7.

Calcular el índice de dificultad del ítem sin corregir y corrigiendo el azar



## II. Discriminación del ítem

- Capacidad del ítem para distinguir o diferenciar a los sujetos de baja, media y alta puntuación en un criterio.
- Se debe conocer tanto en pruebas cognitivas como en las que no lo son.
- El criterio puede ser interno (la puntuación en el test) o externo (otro test u otra variable)
- Cuando el **criterio es interno** el índice de discriminación se puede interpretar como **índice de homogeneidad  $I_H$**  porque expresa el grado de semejanza entre las respuesta al ítems y el resto de los ítems mediados a través de la puntuación total en el test
- Cuando el **criterio es externo** el índice de discriminación es una medida del grado de validez del ítem referido a ese criterio y su valor es proporcional al **índice de validez  $I_V$**  del ítem





- Cuando todas las personas aciertan el ítem  $ID = 1$  o cuando nadie lo acierta  $ID = 0$ , el ítem no es capaz de discriminar.
- La máxima discriminación tiene lugar cuando  $ID = 0.5$  y es acertado por el 50% de las personas.
- Para su cálculo de la Discriminación de un ítem disponemos de dos índices
  - \* Índice de discriminación
  - \* La correlación entre el ítem y el resto del test.
- ***Numero de Discriminaciones = Aciertos (N – Aciertos)***



# Índice de discriminación: Métodos de calculo

## 1. Basada en la comparación del rendimiento de los dos grupos extremos (bajo y alto)

- Se forma los grupos extremos de baja y alta puntuación con el 27% (u otro) de los sujetos en cada uno y luego se calcula  $D$

$$D = \frac{A_{alto} - A_{bajo}}{N}$$

- Si  $D=0$  el ítems no discrimina adecuadamente entre los sujetos de diferentes nivel de aptitud
- Si  $D$  es **negativo** el ítem no contribuye a medir al variable de interés y además favorece la disminución de la precisión del instrumento de medida
  - \* Si  $D \geq 0.40$  el ítem tiene gran poder discriminativo
  - \* Si  $0.30 \leq D \leq 0.39$  la discriminación es aceptable
  - \* Si  $0.20 \leq D \leq 0.29$  el ítem discrimina poco y necesita una revisión
  - \* Si  $0.20 \leq D \leq 0.29$  el ítem no es adecuado debe ser modificado o eliminado
  - \* Si  $D \leq 0.000$  el ítem se debe eliminar directamente



## Ejemplo

Un elemento de un test es contestado por 400 sujetos. Lo contestan correctamente 190; de ellos, 100 pertenecen al 27% de los que mejor puntuación obtienen en el test y 70 al 27% de los que dan un rendimiento más bajo en el test.

Calcular:

1. El índice de dificultad del elemento.
2. El índice discriminativo.
3. El número máximo
4. Discriminaciones que el elemento realiza en esta muestra

Las respuestas de los sujetos al ítem 15 de un test, se distribuyeron según la tabla. La respuesta correcta es la D y tanto las puntuaciones en el test como en el elemento se distribuyen normalmente. Averiguar el índice de discriminación del elemento 15.

Subgrupo	Alternativas del Ítem				
	A	B	C	D*	E
<b>27% Superior</b>	12	10	0	120	12
<b>27% Inferior</b>	12	10	0	40	82



## 2. A través de la correlación entre las puntuaciones en el ítem y en el test.

- Recibe el nombre de **Índice de Homogeneidad** (correlación interna)

		TESTS		
		CONTINUO	DICOTÓMICO	DICOTOMIZADO
ÍTEM	CONTINUO	$r_{xy}$	$r_{bp}$	$r_b$
	DICOTÓMICO		$\emptyset$	$\emptyset_b$
	DICOTOMIZADO			$r_t$

- Se debe descontar de la puntuación total la puntuación del ítem para evitar correlaciones espúreas, o aplicar la siguiente formula

$$R_{j(x-j)} = \frac{R_{jx} S_x - S_j}{\sqrt{S_x^2 + S_j^2 - 2R_{jx} S_x S_j}}$$



## Ejemplos

- Se le aplica una escala tipo Likert a un grupo de 1000 sujetos. La varianza de la escala es de 49 y el ítem 25 tiene una varianza de 2. La correlación del ítem con el test es de 0,40.  
¿Cuál será el índice de homogeneidad de dicho ítem?
- Se construye un test de inteligencia. Se aplica a un grupo normativo pertinente. El test se distribuye según la curva normal. Los elementos son dicotómicos. Se desea conocer el índice de homogeneidad del elemento 25 del test. La media del test en nuestra muestra es de 60 y la desviación típica de 12. La media en el test de los sujetos que han contestado correctamente a nuestro ítem es de 70. El elemento es acertado por el 45 % de los sujetos que responden al test. La varianza del elemento es de 0,45. ¿cuál es el índice de homogeneidad del elemento?

SUJETOS	ELEMENTOS DEL TEST			
	1	2	3	4
A	1	1	0	1
B	1	1	1	0
C	1	1	0	0
D	0	0	1	1
E	0	0	0	0
F	1	1	1	1

Un grupo de 6 sujetos obtuvo en un test de 4 elementos las puntuaciones que aparecen en la tabla adjunta.

Calcular

- El índice de dificultad de los ítems
- El índice de homogeneidad del ítem



### 3. Discriminación mediante contraste de hipótesis

- Queremos ver si la media en el ítem del grupo de sujetos con puntuaciones más alta en el test (el 25% o 27% superior) es significativamente superior a la media en el ítem de sujetos con puntuaciones más bajas (25% o 27% inferior)
- Se utiliza para calcular la discriminación en **test de actitudes**
- Hipótesis
  - Ho: El ítem No Discrimina
  - H1: El ítem Discrimina
- Estadístico de Contraste

$$T = \frac{\bar{X}_{sj} - \bar{X}_{ij}}{\sqrt{\frac{(n_s - 1) S_{sj}^2 + (n_i - 1) S_{ij}^2}{n_{sj} + n_{ij} - 2} \left( \frac{1}{n_{sj}} + \frac{1}{n_{ij}} \right)}}$$



### **Ejemplo**

Se quiere estudiar cierta actitud de los ejecutivos españoles. Para ello se ha elaborado una escala tipo Likert. Se ha aplicado a una muestra de 300 sujetos para hacer un estudio previo del poder discriminativo de los ítems. La puntuación media del 25% con calificación más alta en el ítem 8 ha obtenido una puntuación media de 6,25 y una varianza de 2,12. El 25% con menor calificación en este ítem ha obtenido una media en el mismo de 2 y una varianza de 2,7.

¿ Es la diferencia entre ambas medias estadísticamente significativa? NC 95%.



### III. Fiabilidad y Validez del Ítem

#### Índice de Fiabilidad $IF$

- Correlación del Ítem con un criterio es interno
- Se considera una medida de la precisión con la que el ítem mide el constructo o dominio de interés
- La **fiabilidad** del test se puede expresar en términos de los  $IF$

$$IF = S_j R_{jx} = S_j D_j$$

$$\alpha = \frac{n}{n-1} \left[ 1 - \frac{\sum S_j^2}{(\sum IF)^2} \right]$$

#### Índice de Validez $IV$

- Se evalúa la relación del ítem con un **criterio externo**

$$IV = S_j r_{jy}$$

- El **coeficiente de validez** de un test se puede expresar a través de los  $IV$  de los ítems

$$r_{xy} = \frac{\sum IV}{\sum IF}$$





## Factores que afectan a la discriminación

- Al aumentar la **discriminación** de los ítems aumenta la fiabilidad del test
- Al aumentar la **fiabilidad de los ítems** aumenta la fiabilidad del test.
- Cuando un ítem tiene varianza cero (acertado o fallado por todos) no tiene poder de discriminación.
- El mayor poder discriminativo de un ítem se producen cuando su  $D = 0.5$
- En test multidimensionales el Índice de Discriminación hay que calcularlo respecto al conjunto de ítems asociados a una misma dimensión o constructo

$$S_x = \sum S_j R_{j(x-j)} = \sum \sqrt{p q} R_{j(x-j)}$$

$$S^2 = p q$$

$$\bar{X} = \sum ID$$

$$S_x^2 = \left( \sum IF \right)^2$$



## Ejemplo

- Sometemos cinco sujetos (A, B, C, D y E) a un test de 4 ítems para medir una variable continua y normal. Las calificaciones en cada uno de los ítems así como en el test, aparecen en el siguiente cuadro junto con las notas obtenidas por los sujetos en el criterio del test empleado. Calcular
  1. Índice de homogeneidad del elemento 1
  2. Índice de validez del elemento 1
  3. El coeficiente de validez del test.
  4. El error típico de estimación del criterio a partir del test.

	ÍTEMS				TEST	CRITERIO
	1	2	3	4		
A	5	6	4	7	21	4
B	6	7	5	8	25	6
C	6	7	6	8	27	7
D	7	8	6	8	29	8
E	5	6	4	5	20	5



## IV. Comparación de las respuestas de los ítems

- Para medir el grado de homogeneidad de dos ítems podemos aplicar el estadístico  $\chi^2$  con 1 grado de libertad
- $H_0$ : El grado de acuerdo u homogeneidad entre los ítems **no es significativo** (los dos ítems no miden la misma variable, uno de ellos o los dos son inadecuados)

$$\chi^2 = \frac{n(ad - cd)^2}{(a + b)(c + d)(b + d)(a + c)}$$

- Cuando la homogeneidad es significativa, se calcula el grado o intensidad de dicha homogeneidad como la proporción de sujetos que contestaron correctamente a ambos ítems (este índice representa la consistencia entre los aciertos)

$$p = \frac{(a + d)}{n}$$

	Aciertos	Errores
Aciertos	a	b
Errores	c	d

Podemos plantear además si es grado de dificultad de estos dos ítems es igual o no, es decir si las diferencias encontradas son lo suficientemente pequeñas como para ser debidas a errores de muestreo

$H_0$ : La ejecución de ambos ítems es igual

$$\chi^2 = \frac{(|b - c| - 1)^2}{b + c}$$



## V. Análisis de Distractores

- Se debe analizar primeramente la alternativa correcta, su redacción, el formato, el contenido
- Un **BUEN** distractores
  - ✓ Debe ser elegido por un mínimo de sujetos, como mínimo por el 10% de los sujetos.
  - ✓ Si esta por encima del 10% **NO** es un buen distractor
  - ✓ Deben tener un porcentaje similar de sujetos que elige cada distractor
  - ✓ Debe discriminar en sentido contrario: La correlación de la opción correcta debe ser positiva y la del distractor negativa
  - ✓ La media del distractor debe ser menor que la media de la alternativa correcta

$$\bar{X}_{distractor} < \bar{X}_{correcta}$$

- ✓ La media del distractor debe ser menor que la media total

$$\bar{X}_{distractor} < \bar{X}_{total}$$



## Ejercicio de examen

En la tabla que se muestra a continuación se muestran las respuestas que dieron una muestra de sujetos a un ítem de un test de aptitudes. La respuesta correcta es la A. A un NC del 95%, ¿ podemos decir que las alternativas incorrectas son equiprobables?: a) sí porque el valor teórico del estadístico es menor que el valor obtenido; b) sí porque el valor teórico del estadístico es mayor que el valor obtenido; c) no porque el valor teórico del estadístico es mayor que el obtenido.

A	B	C
389	52	36



## VI. Funcionamiento Diferencial del Ítem

Analiza si distintos grupos con el mismo nivel de competencia tienen distintas probabilidades de éxito en un ítem, pero sin apuntar a posibles causas

### ❖ Sesgo

- Es una fuente de invalidez o error sistemático y contrario al error aleatorio
- Son diferencias encontradas entre distintos grupos medidos con el mismo instrumento de medida, estas diferencias se pueden deber a diferencias reales entre los grupos o a fuentes sistemáticas de variación ajenas al constructo
- Un ítem está **sesgado** cuando la probabilidad de éxito para sujetos igualmente competentes en la variable que mide el **ítem depende del grupo al que pertenezca el sujeto**
- Un ítem es **insesgado** si la probabilidad de éxito para el ítem es la misma para sujetos igualmente capaces **independientemente del grupo al que pertenezca.**

### ❖ Impacto

- Un ítem presenta impacto cuando refleja una **diferencia real (NO UN ERROR SISTEMÁTICO)** entre grupos en la variable medida por el ítem debido que la probabilidad de responderlo correctamente es mayor para un grupo que para otro.
- Implica diferencias reales entre los sujetos de ambos grupos

### ❖ FDI

- Cuando existen diferencias en la puntuación media obtenida en ese ítem por dos grupos distintos de sujetos pero con el mismo nivel en el rasgo
- Indica que las diferencias no son reales



## Ejercicios libro

**199.** En la tabla siguiente aparecen las frecuencias de sujetos, ordenados en 5 percentiles según su puntuación, que respondieron a un ítem de un test de rendimiento en lengua con 4 opciones de respuesta, donde la opción C es la correcta. La media y varianza en el test de los sujetos fue de 6 y 4 respectivamente, mientras que la media en el test de los sujetos que acertaron el ítem fue de 7.

	A	B	C	D
20% superior	8	17	207	11
20% siguientes	11	39	223	25
20% siguientes	16	53	183	32
20% siguientes	17	61	113	29
20% inferior	43	70	106	39
<i>P</i>	0,07	0,18	0,64	0,10

Calcular:

1. El índice de dificultad del ítem sin corregir y corrigiendo el azar.
2. El índice de discriminación del ítem.



**212.** Aplicamos un test de inteligencia a un grupo de 300 sujetos. La respuesta se exige en el formato «verdadero-falso». El elemento n.º 16 del test es contestado por todos los sujetos, pero sólo 180 lo hacen de forma correcta. La  $\bar{X}$  en el test de los sujetos que responden correctamente el ítem 16 es de 70 puntos. La media obtenida por el grupo en el test es de 65 puntos y la varianza de 256. La varianza del elemento n.º 16 es de 0,24. ¿Cuál es el índice de discriminación o de homogeneidad de dicho elemento?

**215.** Se ha construido un test que se corrige clasificando a los sujetos en dos categorías: aptos-no aptos. Todos los elementos del test son dicotómicos. Se aplica el test a un grupo normativo y deseamos calcular el índice de homogeneidad de uno de sus elementos. Los sujetos han contestado según muestra la tabla adjunta.

	TEST	
	Aptos	No aptos
Acierta	120	30
Falla	20	150





**216.** Hemos construido un test de rendimiento académico. El test sigue una distribución normal. Sus elementos aunque normales y continuos están dicotomizados. Deseamos calcular el índice de homogeneidad del elemento 4. La media del test es de 50 puntos y la desviación típica de 15. La media en el test de los sujetos que han acertado el elemento 4 es de 55. La varianza del elemento es 0,81 y la proporción  $P/Y$  es igual a 1,7549. ¿Cuál es el índice de homogeneidad de dicho elemento?

**227.** A un grupo de 500 sujetos se les ha aplicado un test de 100 elementos. Cada elemento tiene 5 alternativas de las cuales sólo una es correcta. Las respuestas dadas al elemento 80 del test son las que figuran a continuación. (La respuesta C es la correcta.)

1. Analizar las alternativas de dicho elemento.
2. Calcular el índice de discriminación.
3. Calcular el número de discriminaciones del elemento.

Subgrupo	Alternativas del ítem					
	A	B	C	D	E	Omisiones
27% superior	5	10	80	20	20	0
27% inferior	25	60	20	5	20	5



**237.** Para medir una determinada actitud se quiere construir una escala siguiendo la técnica de Likert. Se escoge una muestra de 60 sujetos a los que se les aplica un conjunto inicial de 30 elementos. Cada uno de estos elementos tiene 5 categorías de respuesta que van del 1 al 5. Ordenados los sujetos según la puntuación total obtenida, seleccionamos el 25% que obtuvo puntuaciones más altas y el 25% de los que obtuvieron puntuaciones más bajas. Las puntuaciones obtenidas por ambos grupos en el elemento n.º 15 de la prueba fueron las siguientes:

Grupo alto 25% superior	1	3	1	4	4	1	2	5	5	5	5	4	4	5	5
Grupo bajo 25% inferior	1	4	1	2	2	1	5	1	2	2	2	1	3	1	2

1. Calcular el poder discriminante del ítem mediante la prueba estadística  $t$  de Student utilizando un nivel de confianza del 95%.
2. Utilizando el test de la mediana y aplicando la prueba de Chi cuadrado, ¿podemos decir que el ítem es discriminativo al NC del 99%?

*Nota.* No tener en cuenta el pequeño tamaño de la muestra.